在线医疗社区用户画像构建与应用*

■ 滕春娥[1] 何春雨[2]

[1] 牡丹江师范学院图书馆 牡丹江 157011 [2] 牡丹江师范学院历史文化学院 牡丹江 157011

摘 要：[目的/意义]在线医疗社区用户画像能够精练化提取患者需求、可视化描述患者特征，有助于平台管理者制定更精准的信息服务策略。[方法/过程]在梳理在线医疗社区用户画像研究成果的基础上，提出一种在线医疗社区用户画像构建框架，并爬取百度贴吧中糖尿病社区用户数据，综合采用档案清单、复合图形、语言描述和词云方法，实现了具有4个维度和12项标签的在线医疗社区用户画像构建。[结果/结论]该理论框架及实现方法有助于信息精准推荐、改善患者体验、提升决策质量，为在线医疗社区用户画像构建与应用提供了理论与实践指导。

关键词：用户画像 在线医疗社区 糖尿病

分类号：G250

DOI：10.13266/j.issn.0252－3116.2021.12.015

1 引言

"互联网＋"背景下涌现出平安好医生、好大夫在线、微医和春雨医生等在线医疗社区（Online Health Community，OHC），它们将传统医疗服务转移到线上环境，并不断重塑诊疗流程，包括医患匹配、医患沟通和个性化医疗服务。实现医生与患者、话题与患者、健康知识与患者等的精准匹配，不仅能够提升用户获取信息的质量和效率，还有助于增强在线医疗社区提供定制服务的有效性。精准匹配需要以条目精细、描述清晰的用户需求为基础，从而据此推送适宜的医生、话题和健康知识。用户画像（User Persona，UP）作为一种用户需求发现、匹配和建模的新兴工具，是面向在线医疗知识管理与服务中的研究重点和热点[1]。

UP是以用户为原型和中心设计（User-Centered Design，UCD）[2]，通过还原属性特征和给予场景描述[3]组建的可视化标签集合。依托在线医疗应用程序（如PC端的网页社区、移动端的APP等）聚集的海量数据，提炼用户属性和构建用户画像，是在线医疗平台精准化推荐和服务的前提[1]。国内的在线医疗社区用户画像（"User Persona in Online Health Community"，简称"OHC-UP"）研究相对较少，除综述性文章外仅有10篇相关文献（引文数据库：CNKI；检索日期：2020年10月31日）。如张海涛等利用Con Exp工具将用户标签换算为概念格，通过概念格Hasse图将用户群体分类，实现健康社区大数据群体画像[4]。国外对OHC-UP研究较为重视，尤其是以患者为中心（Patient-Centered Design，PCD）的用户画像成为研究热点，已建立基于PCD的用户画像包括面向心血管疾病药物治疗者[5]、传统药用植物使用者[6]、慢性病老年人[7]等；工作、性别、年龄、IT经验、网站搜索能力、身体特征、认知特征、工作特征、生活方式、社会支持、移动端使用、使用焦虑、健康信息搜索等精细化标签已成功应用于患者画像[5,7-8]。这些患者画像的应用场景包括电子健康偏好判断、可视化诊疗档案、个人健康状况评估等[5-8]。但是，上述研究对OHC-UP的构建方案缺乏一般性，对具体场景中应用方式的解读还有待丰富。

基于此，本文综合采用多种标签表达和画像生成方法，提出一种更具一般性的OHC-UP构建框架，将其应用于实践来增强研究的落地效力和实用价值，并明确OHC-UP的应用能够解决何种现实问题，以期为完善在线医疗平台的知识服务提供参考和借鉴。

2 在线医疗社区用户画像构建框架

OHC是为用户提供医疗信息服务的网络平台，本

* 本文系黑龙江省哲学社会科学规划项目"基于大数据分析的以满足个性化需求的精准出版服务研究"（项目编号：18XWD374）研究成果之一。

作者简介：滕春娥（0000－0003－3721－913X），副研究馆员，博士；何春雨（0000－0001－8839－1445），副教授，博士，通讯作者，E-mail：teng0128fei@163.com。

收稿日期：2020－03－10 修回日期：2021－05－28 本文起止页码：147－154 本文责任编辑：易飞

文主要讨论面向患有或可能患有某些疾病人群的 OHC-UP。在借鉴现有文献思路的基础上[5-8]，提出一种具有一般性的 OHC-UP 构建框架，框架构建核心步骤包括标签体系建立、社区数据采集、用户画像呈现。如图 1 所示：



图 1　在线医疗社区用户画像构建框架

2.1 标签体系建立

建立 OHC-UP 标签体系的主要目的是创建一种基于 PCD 的还原性患者角色，作用是以细分的标签来挖掘 OHC 患者的需求、愿望和障碍。基于 PCD 的标签建立过程的基本原则包括：①将患者原型置于设计中心；②优先关注患者及其任务；③确保可用性。基于上述原则，本文自上而下地建立 OHC-UP 标签体系，包括标签顶层设计、标签凝练细化、标签表达规则三个环节。

2.1.1 标签顶层设计

依据将患者原型置于设计中心的原则，标签顶层设计需要以不同侧面的、加以区别的和全方位的视角还原研究对象。不同于现有研究仅关注某个侧面[4-8]，本文总结并集成了 OHC 用户的 4 个核心属性：①自然人属性。具有确定的人口学统计特征和自然人天然具备的一切需要[7-8]。②患者人属性。信息服务对象是患有或可能患有某些疾病的人群，需要根据发病部位、种类和对应科室等因素进行功能性服务细分[6-7]。③在线人属性。OHC 赋予用户虚拟的角色和身份，允许用户展示与其自然人真实身份相区别的网络形象[4]。④社区人属性。具有维系群体关系的需要，自发产生一种对社区群组的依恋感，建立与其他用户的人际纽带[4]。如图 2 所示：

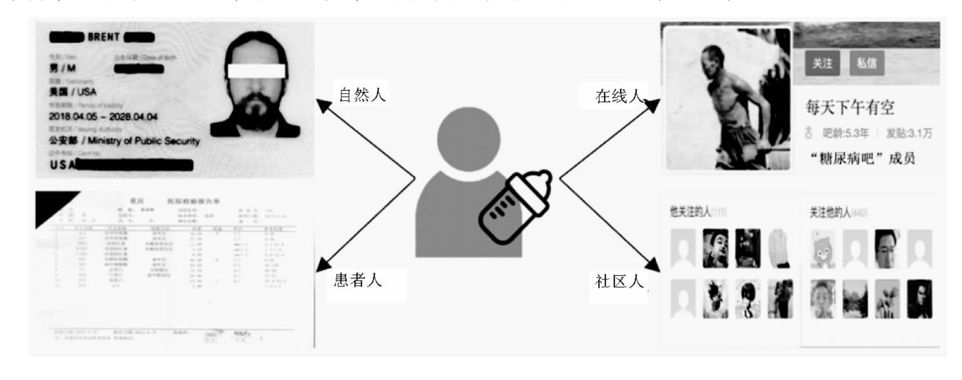


图 2　在线医疗社区用户核心属性细分

2.1.2 标签凝练细化

依据优先关注患者及其任务原则，所选取标签应当具有 PCD 职能，即为用户提供帮助、传达用户需求、用于任务分析、用于客户服务。基于此，本文参考现有文献的标签设计，细化前文 4 个核心属性的下设标签。见表 1。

（1）自然人属性。姓名（或身份证号）是 OHC 用户在真实世界中的基本标识，用于社区管理者和其他成员识别和确认他人真实身份[7]；年龄、性别和居住地是基于人口学特征细分用户的常用标准，可作为信息推送（如医院信息、医生信息、日常养护建议等）的基础参考依据[4]；职业旨在描述 OHC 用户在现实生活中以何种方式获得收入以及需要为此付出何种劳动，便于社区和医务人员针对性地提供信息服务[7-8]；愿望是 OHC 用户对现实生活最为期待的诉求，用于社区管理者或医务人员为患者提供核心关怀[7]。

（2）患者人属性。病种是 PCD 设计者分类 OHC 用户最重要的标签，用于对医疗健康信息服务作出功

能性区分[5,7-8];患病时间、疾病危险性是判断生理(或心理)健康需求紧迫性的基本依据,被看作 OHC 用户健康状况的描述[7-8]。

(3)在线人属性。昵称(或用户 ID)是 OHC 用户在虚拟社区中的基本标识,用于社区管理者和其他成员识别和确认他人虚拟身份[4];信息负载是用户参与 OHC 时需要接触信息量的大小,用于表示 OHC 用户主观上愿意投入于健康信息服务的时间和精力[9];信息偏好是 OHC 用户对社区内各类健康信息的喜好程度,用于支持推送系统针对个人偏好开展精准信息荐阅[4,10]。

(4)社区人属性。社交类型旨在刻画 OHC 用户的社交意愿、需求和能力,可辅助社区决定是否应向用户推送社交信息及信息量的多少[10];社会情感旨在识别用户向 OHC 其他成员传递出何种情感,是积极、消极抑或是中立[11]。

表 1 在线医疗社区用户画像标签体系

顶层维度	具体标签
自然人属性	姓名、年龄、性别、居住地、职业、愿望
患者人属性	病种、患病时间、疾病危险性
在线人属性	昵称、信息负载、信息偏好
社区人属性	社交类型、社会情感

2.1.3 标签表达规则

根据确保可用性的原则,需要明确表 1 中 OHC 用户标签的具体表达方法和预期结果。部分标签采用自然语言描述法进行表达,直接从证据性文本中抽取内容,见表 2。下面着重说明标签的表达规则:

(1)病种。为取得统一、规范的病种编码与名称,将国际疾病分类(International Classification of Diseases, ICD)导入 Excel 数据库[12],作为模糊匹配用户 UGC 中对患病种类描述的参照。

(2)疾病危险性。病情发展的严重程度决定病人就诊与处置的优先次序,如将病人按照濒危、危重、急症、非急症进行分流,根据病种按照低危、中危、高危、极高危进行分级等。

(3)信息偏好。根据现有文献的建议,可通过用户主页关注的社区话题抽取关键内容,或者提取用户对话数据中的高频词,来确定个人信息偏好[4]。

(4)社交类型。现有文献根据 OHC 用户社交网络的出入度相对高低进行社交类型区分,认为若出入度数值分布符合正态分布,应以 3σ 原理作为判断依据,反之采用二八定律为区分准则[13]。本文基于网络节点出入度的用户社交类型矩阵,将用户社交类型划分为达人型、贵族型、粉丝型和佛系型 4 种。如图 3 所示:

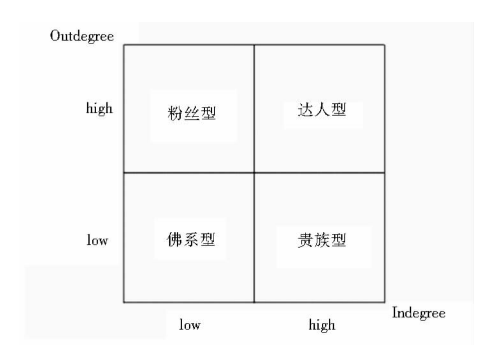


图 3 基于网络节点出入度的用户社交类型矩阵

(5)社会情感。朴素贝叶斯是广为采纳的情感分析方法,通过语料库训练情感分类模型,判断患者社区对话过程中隐含的情绪状态[14]。

表 2 在线医疗社区用户画像标签表达规则

标签	表达方法	表达结果
姓名	自然语言描述法	如"高树新"
年龄	自然语言描述法	如"63 岁"
性别	自然语言描述法	男/女
居住地	自然语言描述法	如"北京"
职业	自然语言描述法	如"演员"
愿望	自然语言描述法	如"能看到我孙女长大和结婚"
病种	基于分类的类型范围(由评论文本/诊断图片确定)	如"霍乱(A00.901)"
患病时间	自然语言描述法	如"3 个月"
疾病危险性	基于分类的类型范围(对疾病发展程度分级)	低危/中危/高危/极高危
昵称	自然语言描述法	如"飞翔的企鹅"
信息负载	评分量表(由平均活跃时长进行分级)	天使/忠实/常规/游客/僵尸
信息偏好	基于统计特征的类型范围(由关注话题/评论文本确定)	如"三鹿奶粉"
社交类型	基于统计特征的类型范围(由网络节点出入度确定)	达人型/贵族型/粉丝型/佛系型
社会情感	评分量表(根据评论文本进行情感分析)	积极/中立/消极

2.2 社区数据采集

社区数据采集是 OHC-UP 构建过程中承上启下的中间步骤,在完成标签体系建立后,需要通过 OHC 数据提取信息,支持 UP 标签表达过程。这里将社区数据采集细分为社区数据爬取、数据清洗加工、画像数据准备三个环节。

2.2.1 社区数据爬取

不同于现有文献优先获取数据,从而挖掘特征、提

炼标签的主张[4-5]，即，本文主张优先确定标签，按照社区管理者需求，收集用于表达特定标签的数据。构建 OHC-UP 需要爬取的社区数据包括用户主页数据、用户对话数据、用户关注数据、用户诊疗数据。诚然，对于 OHC 管理者则无需爬取社区数据，仅需从服务系统的后台数据库中完成调取工作。但是，厘清上述数据的描述内容、应用方法与 UP 标签的逻辑关系仍然是有益的。如表 3 所示：

表 3 在线医疗社区爬取用户数据描述

类型	主要形式	可信度	应用方法	服务标签
用户主页数据	文本/图片	相对较低	提取文本	姓名、年龄、性别、居住地、职业、愿望、昵称、信息负载、信息偏好
用户对话数据	文本	准确	词频分析、情感分析	信息偏好、社会情感
用户关注数据	文本	准确	社会网络分析	社交类型
用户诊疗数据	图片	相对较高	提取文本	姓名、年龄、性别、居住地、病种、患病时间、疾病危险性

2.2.2 数据清洗加工

OHC 数据清洗和加工具体包括[13]：①剔除无效值；②重复项合并；③数据规范化；④关键内容抽取。经过上述数据处理工作后，将数据存储至合适的数据库（如 SQL Server），形成初步数据集。

2.2.3 画像数据准备

形成画像数据集合，还需要根据实际研究情况，剔除初步数据集中含有 NULL 值、可信度较低的低质量数据，确保 OHC-UP 的任意项标签都有相应的、可信的支持数据，理由如下：①不完全数据直接限制了标签表达过程，使最终的结果出现“残缺”画像；②现有研究发现用户填写的个人信息、生成的网络评论很可能存在与事实不一致的情况[15]。因此，以不完全数据或低可信度数据构建 OHC-UP，均会降低画像的决策支持能力和社区管理者的决策质量。故本文所用的画像数据，本质上是为满足研究者所追求的适宜粒度，过滤大数据得到的规模被缩减、信度被提高的优质数据集。

2.3 用户画像呈现

对于 OHC-UP 生成，本质是运用预先准备的数据集，按照既定标签体系中的标签类别、条目及其表达规则，逐类逐项地计算患者原型的标签表达结果，赋予标签信息，并在完成全部计算、“贴满”标签后，以可视化、便于应用乃至别出心裁的画像建立最终结果，为 OHC 管理者或医务工作者提供借鉴和参考。

本文构建了具有 4 个维度和 14 项标签的在线医疗社区用户画像，其中 8 项标签采用自然语言描述法、2 项标签采用评分量表、4 项标签采用类型范围描述。该画像的底层支持技术包括数据挖掘、社会网络分析和情感分析，呈现方式综合采用前沿理论和广泛应用的档案清单、复合图形、语言描述和词云 4 种途径。

3 在线医疗社区用户画像应用

本文依循上述构建的 OHC-UP 框架，遵循第 2 部分的理论步骤开展实证检验，利用百度贴吧生成糖尿病吧社区的用户画像，进一步说明该理论框架的可行性和应用价值。

3.1 标签选择

根据百度贴吧糖尿病吧社区用户的数据特征，对 2.1 节的一般性 OHC-UP 标签体系进行修正：①由于无法征求被试者的隐私意愿，姓名标签因科研伦理限制被删除；②由于贴吧用户的职业为非必填信息项，职业标签因数据采集限制被删除。

最终选取昵称、年龄、性别、居住地、愿望、病种、患病时间、疾病危险性、信息负载、信息偏好、社交类型、社会情感共 12 项标签。结合糖尿病吧的具体社区特征，对标签表达规则作出如下修正：①居住地推断为就诊医院所在地；②患病时间推断为诊断报告时间；③疾病危险性参考患者描述、化验指标和辅助报告；④愿望缺乏证据性文本时以自动摘要为代理变量[16]；⑤以平均活跃时长作为信息负载的代理变量（平均活跃时长为发帖数与吧龄的比值），信息负载以 0.20 为等距区间分为 5 个等级。

3.2 数据采集

本文通过 python 编写爬虫代码，爬取 2020 年百度贴吧的“糖尿病吧”主题帖中参与用户的主页数据和发帖数据。选择百度贴吧中“糖尿病吧”展开研究的具体理由包括：①拥有较大规模的用户作为研究对象[9]；②已积累丰富的患者生成内容和交互数据[9]；③能够获取患者公开的诊疗材料的图片；④糖尿病列入中国十大高发疾病排名[17]；⑤用户信息行为数据具有动态性和时效性[11]。

通过主页数据可获取参与用户的吧龄、发帖数、关注帖、关注列表、被关注列表，用于计算信息负载、信息偏好、社交类型共 3 项标签。见图 4。发帖数据包括评论文本（对话数据）和诊断图片（诊疗数据）。评论文本作为直接性证据、生成自动摘要或情感分析的语料，

用于计算昵称、愿望、社会情感共 3 项标签。见图 5。识别并提取诊疗图片中的关键内容，用于计算年龄、性别、居住地、病种、患病时间、疾病危险性共 6 项标签。见图 6。按照 2.2.2 和 2.2.3 节的方法对上述数据作进一步处理，形成含有 85 条数据的画像数据集。见图 7。

这里对病种、信息负载、社交类型的标签锚点作必要说明：本文是首篇将国际通用的疾病编码类目引入 OHC-UP 研究的文献，糖尿病（E10-E14）包括胰岛素依赖型、非胰岛素依赖型、营养不良相关性、其他特指、未特指 5 个类目[12]，操作性标准包括年龄、体型、酮症酸中毒、胰岛素治疗效果、其他指标和疾病报告等，故本文虽以糖尿病社区用户为数据来源，被试者的病种却并不是完全同质的。其中，E10 为胰岛素依赖型糖尿病，E11 为非胰岛素依赖型糖尿病，E12 为营养不良相关性糖尿病，E13 为其他特指的糖尿病，E14 为未特指的糖尿病。

根据画像数据计算全体用户的信息负载指标，降序确定锚点呱呱皮大＊（Order = 17，发帖数为 2 817，吧龄为 6.5）、破昵称真难＊（Order = 34，发帖数为 1 197，吧龄为 8.6）、王老实我没＊（Order = 51，发帖数为 196，吧龄为 4）、清椅白＊（Order = 68，发帖数为 124，吧龄为 7.9）、请叫我刘总＊（Order = 85，发帖数为 3，吧龄为 6.7），得出天使、忠实、常规、游客、僵尸用户的所属区间依次为（＋∞，433.38］、（433.38，139.19］、（139.19，49.00］、（49.00，15.70］和（15.70，0.00］。

按照文献要求的划分方法，计算本文用户社交类型的所属区间[13]。当 $F(x \leq x_0) < 0.8$ 时，选取 x_0 作为分界线，确定入度锚点为迪卡农＊（入度为 116，$F(x \leq 116) < 0.8$）和的记＊（入度为 119，$F(x \geq 119) < 0.2$），出度锚点为的记＊（出度为 43，$F(x \leq 43) < 0.8$）和 fyudlts007＊（出度为 44，$F(x \geq 44) < 0.2$），得出 in-degree > 116 且 out-degree > 43 为达人型、in-degree > 116 且 out-degree < 44 为贵族型、in-degree < 119 且 out-degree > 43 为粉丝型、in-degree < 119 且 out-degree < 44 为佛系型。

3.3 画像呈现

本文创建 OHC-UP 应用设计的模型实例，提供了一种编码和识别用户配置和角色属性的信息系统。限于篇幅，本文仅报告两则实例，见图 8 和图 9。不同于现有 OHC-UP 文献的采纳标签和呈现设计[10,20]，本文补充了公众的内部心理力量、外部社会力量对医疗保健选择和结果的影响[24]。这使本文扩展了传统的人物资料和个体属性中未曾设计的角色参考，包括：①愿望，通过基于句子的文本生成摘要体现 PCD 对利益相关者的核心关怀[16]；②疾病/糖尿病检测和诊断，根据患者提供的医学报告提取病种、病情等信息，并利用 ICD 编码将疾病/糖尿病的类型具体化；③信息服务，通过词频统计掌握并排序用户有限的兴趣爱好和关注议题；④社会支持网络，对主页的关注数据进行社会网络分析构建患者网络，明确不同患者差异化的社交需求。



图 4 “糖尿病吧”用户主页数据



图 5 “糖尿病吧”用户评论文本

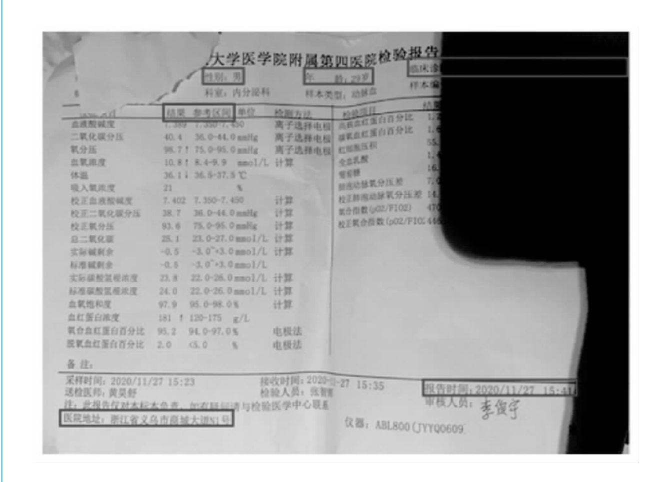


图 6 “糖尿病吧”用户诊断图片



图 7 “糖尿病吧”画像数据集示例



图 8 被试者“如蠓如惑 * ”OHC-UP 实例



图 9 被试者“Shengxing * ”OHC-UP 实例

在 OHC-UP 应用过程中，可视化技术增强了患者原型和角色特征的呈现效果。整体设计采用 C. Lerouge 等推崇的档案清单[7]，学者认为档案清单是在公众普及医疗保健技术背景下，被管理者广泛接纳的捕捉、设计和开发用户特征的增强型人物工具；印象设计采用 V. Haldane 等推崇的复合图形[5]，以适合阅读者认知结构、易于理解与推测、快速捕捉关键特征与细节、利于养成短期记忆的图形组合呈现 OHC 用户社交类型等重要标签。通过语言描述和词云辅以常规设计，语言描述是 UP 呈现模式中最直接的信息反馈手段，它将凝练后的关键特征和证据性文本，利用平铺直叙、最具说服力的方式予以展示[7]，而词云可更好地体现用户信息偏好中的主题范围及其位序[3]。

相比于现有研究[3-5]，本文在画像标签上扩至 12 项，在生成方式上扩展至 4 种。本文 OHC-UP 较好的呈现效果还得益于优质的画像数据集。现有决策理论表明，当决策条件繁多而复杂时，决策者即使付出更多的努力也未必能得出最优决策[11]。某种程度上，基于本文数据的画像由于粒度和复杂程度均相对适中，具有更强的决策支持能力。若将画像数据集扩展为原始数据集，并不影响本文 OHC-UP 应用设计的实现，而只是在生成结果中出现含有空白标签的实例。

4 结果分析和启示

OHC-UP 的成功实施和推广，很大程度上取决于社区服务提供者的认知[7,18]。这意味着要使患者成功地对社区及其信息服务产生依赖，开发人员必须认识到，采用 OHC-UP 的关键是服务提供者和服务对象的接受，即采用 OHC-UP 应有应用价值，患者有愿望让开发人员使用该技术提供信息服务[7]。故本文结合生成的 OHC-UP 实例，详细阐释如何通过 OHC-UP 场景应用，满足信息精准推荐、改善患者体验、提升决策质量的社区实践需求。

4.1 信息精准推荐

两则 OHC-UP 实例中，“如蠓如惑 * ”是一例低危型营养不良相关性糖尿病患者、高信息负载水平的天使用户，“Shengxing * ”则是一例高危型胰岛素依赖型患者、中信息负载水平的常规用户。若为上述用户荐阅参考性就诊信息和专业性康养知识，在信息类型和内容上，前者需要关注主题应聚焦于引发其糖尿病的其他疾病，建议其就近就医，而后者更适合获悉有效降糖、饮食控制、运动计划等诊疗方案，应考虑在权威医疗系统及早就诊；在信息体量和频率上，前者更适合提供高信息体量和频率的推送服务，而后者却可能会对繁多的信息和推送产生厌倦情绪。

相比于现有画像文献的推荐模式[4-5]，本文提供了更加丰富的标签（病种、疾病危险性、信息负载等），

从而服务于信息精准推荐。从信息角度来看,OHC 信息服务被要求给予更多的个性化关注,且技术工具的使用有助于节省其捉襟见肘的资源(如时间)[7,9,11]。本文的 OHC-UP 作为一种信息系统,它输出的用户档案清单能够高效地呈现出用户必要的个性化关注和反馈。同样地,这也阐明了 OHC-UP 如何能够并且应该作为社区管理者为患者提供精准信息服务的重要工具。

4.2 改善患者体验

两则 OHC-UP 实例中,"如蓁如惑 * "是一位爱好社交、情绪低落的青年男性,伴有痛风、焦虑症和艾滋病等疾病,社区应为其荐阅饲养鸟类宠物的有趣资讯、其他病友成功对抗疾病的治疗案例等信息,引导其与乐观的社区成员发展有益的互动关系;"Shengxing * "则是一位淡泊社交、情绪中立的中年男性,是爱好徒步、养植、诗词、收藏的发烧友,应通过社区服务放大其徒步兴趣,并及时通知近期关于植物、诗词和收藏的展览会等线下活动,尽可能增加该患者的运动需求和外出意愿。

相比于现有画像文献的服务理念[4-5],本文更加重视对患者的人本关怀,特别是不同性别、年龄段和并发症等患者的社交与情绪状况。从患者角度来看,患者其实并不缺乏糖尿病意识和知识的资源[7],而是难以偶遇有益于调节身心健康和具体疾病的集中和感兴趣的信号和机会。即患者可以获得多种其他信息源,但是信息源的相对质量和可靠性有待考证[15,19],而且发现的信号和机会与自身疾病基本没有相关性。故除用户的自然人和患者人属性之外,本文的 OHC-UP 还关注了在线人和社交人属性,通过多维度属性融合的关联特征挖掘,为患者提供与自身疾病非直接相关却有益的信息服务,激活其社交驱动的兴趣和保健行为。

4.3 提高决策质量

两则 OHC-UP 实例中,二者的愿望有较大区别,"如蓁如惑 * "陈述了疾病对个人心理状态带来的巨大困扰,"Shengxing * "则表达个人对健康饮食的向往和对非健康饮食的警诫。事实上,社区管理者能够根据用户愿望来推测其深层次的心理特征[7],故前者其实是更渴望 OHC 服务给予更多来自社会的正外部性(如鼓励、安慰和频繁的互动等),而后者看似是致力于饮食计划的改善,本质上却是对个人兴趣的扩展与激活(如对诗词、藤蔓植物、奇石和陨石的欣赏等)。

相比于现有画像文献的决策思维[4-5],本文不同于基于患者心理与行为特征挖掘、用户习惯决定论的社区运作决策,而是选择了基于患者获取服务目标的核心愿望挖掘、用户初衷决定论的社区运作决策。个人偏好具有一定的复杂性[3,9],管理者基于这种考量可能产生不止一个片面性的错误认知。考虑个人偏好的服务策略有时并不是有效的,只有通过"walking in their shoes"(一种西方学者的隐喻),真正为用户设计,才能化解决策团队意见冲突[7]。本文 OHC-UP 在充分捕捉用户特征的基础上,重视用户渴望的核心关怀,有助于决策者对用户的心智产生交换性认知。

5 结语

在线医疗社区是互联网赋能医疗服务业而产生的新业态,实践界迫切地需要学者设计和开发辅助技术,更好地协助社区管理者为患者提供信息服务。在理论层面上,本文析出了 OHC-UP 构建框架,从标签体系建立、社区数据采集、用户画像呈现三个核心步骤,阐释了 OHC-UP 构建过程的底层逻辑与实现机制,并细化出具体的、一般性、可复制的操作环节,该理论框架为在线医疗社区用户画像构建与应用提供了理论依据和方法指导。在实践层面上,本文爬取百度贴吧中糖尿病社区用户数据,综合采用档案清单、复合图形、语言描述和词云,实现了具有 4 个维度和 12 项标签的在线医疗社区用户画像,论述了本文 OHC-UP 在构建框架和实现方法的可行性和应用价值,以及它能应用于何种场景、解决何种问题。

以下内容对弥补本文局限具有重要价值,包括:①OHC 服务对象是具有健康信息需求的人群,多以患者为核心客户和流量来源[9,19]。本文对上述社区具有较好的适用性和应用价值,却不适用于以医生为核心客户的 OHC,未来应针对较小众的在线医生社区开展 UP 研究;②实证部分未能排除 A. J. McLeod 和 J. G. Clark 等学者的担忧,即 OHC 实际操作系统用户可能与假定为主要用户的人不同[20],可能存在家属、公益人士等代为生成社区内容(如帖子、诊断图片)的情况,需要研究者提供适合的识别技术和算法;③未来应对病种标签的 ICD 编码作精细化处理,扩展为疾病类目的亚目、细目高阶分类的四、五位数编码[12],提供更具有针对性的诊疗保健服务。

参考文献:

[1] 马费成,周利琴.面向智慧健康的知识管理与服务[J].中国图书馆学报,2018,44(5):4-19.

[2] MASSANARI A L. Designing for imaginary friends: information architecture, personas and the politics of user-centered design[J].

New media & society, 2010,12(3):401 – 416.
[3] 宋美琦,陈烨,张瑞. 用户画像研究述评[J]. 情报科学,2019,37(4):171 – 177.
[4] 张海涛,崔阳,王丹,等. 基于概念格的在线健康社区用户画像研究[J]. 情报学报,2018,37(9):912 – 922.
[5] HALDANE V, KOH J J K, SRIVASTAVA A, et al. User preferences and persona design for an mhealth intervention to support adherence to cardiovascular disease medication in singapore: a multi-method study [J]. JMIR mhealth and uhealth, 2019, 7(5): e10465.
[6] ISA W A R W M, AMIN I M, ISHAK N. Designing mobile information architecture (ia) m-health learning application for traditional malay medicinal plants with medicinal properties using user persona [J]. Advanced science letters, 2018, 24(1):603 – 607.
[7] LEROUGE C, MA J, SNEHA S, et al. User profiles and personas in the design and development of consumer health technologies[J]. International journal of medical informatics, 2013, 82(11):251 – 268.
[8] TURNER A M, REEDER B, RAMEY J. Scenarios, personas and user stories: user-centered evidence-based design representations of communicable disease investigations[J]. Journal of biomedical informatics, 2013, 21(4):575 – 584.
[9] 尚丽维,郭勇,张向先. 在线医疗社区信息交互关系网络关键节点影响力机理研究[J]. 情报理论与实践,2020,43(8):140 – 145.
[10] GEVA H, OESTREICHER-SINGER G, SAAR-TSECHANSKY A M. Using retweets when shaping our online persona: topic modeling approach [J]. MIS quarterly, 2019.43(2):501 – 524.
[11] 郭勇,高歌,王天勇,等. 社交网络舆情意见领袖研究:蝴蝶图示、甄别及影响力评价[J]. 图书情报工作,2019,63(14):62 – 73.
[12] WORLD HEALTH ORGANIZATION. International classification of diseases 11th revision[EB/OL]. [2020 – 09 – 01]. https://icd.who.int/en.
[13] 陈烨,王乐,陈天雨,等. 基于社会网络分析的社会化问答平台用户画像研究[J]. 情报学报,2021,40(4):414 – 423.
[14] 周红磊,张海涛,张鑫蕊,等. 话题 – 情感图谱:突发公共卫生事件舆情引导的切入点[J]. 情报科学,2020,38(7):15 – 21.
[15] 国佳,郭勇,沈旺,等. 基于在线评论的网络社区信息可信度评价方法研究[J]. 图书情报工作,2019,63(17):137 – 144.
[16] 唐晓波,顾娜,谭明亮. 基于句子主题发现的中文多文档自动摘要研究[J]. 情报科学,2020,38(3):11 – 16,28.
[17] 中国报告大厅. 中国十大高发疾病调查排名[EB/OL]. [2021 – 06 – 01]. http://www.360doc.com/content/19/1021/11/5316345_868130093.shtml.
[18] GEFEN D, STRAUB K D W. Trust and TAM in online shopping: an integrated model[J]. MIS quarterly, 2003, 27(1):51 – 90.
[19] 郭勇,张海涛. 新冠疫情与情报智慧:突发公共卫生事件疾控应急工作情报能力评价[J]. 情报科学,2020,38(3):129 – 136.
[20] MCLEOD A J, CLARK J G. Using stakeholder analysis to identify users in healthcare information systems research: who is the real user? [J]. International journal of healthcare information systems & informatics, 2009, 4(3):1 – 15.

作者贡献说明:
滕春娥:提出研究方案,撰写论文,处理与分析数据;
何春雨:设计研究框架,提出修改意见,修订论文。

Construction and Application of User Persona in Online Health Community

Teng Chune[1] He Chunyu[2]
[1] Library of Mudanjiang Normal University, Mudanjiang 157011
[2] College of History and Culture, Mudanjiang Normal University, Mudanjiang 157011

Abstract: [**Purpose/significance**] Online health community (OHC) user persona (UP) can concisely extract the needs of patients, visually describe the characteristics of patients, and help platform managers to formulate more accurate information service strategies. [**Method/process**] On the basis of combing OHC-UP research results, this paper proposed a construction framework of OHC-UP, crawled the user data of diabetes community in Baidu Post Bar, and comprehensively used file list, composite graphics, language description and word cloud to realize OHC-UP with 4 dimensions and 12 labels. [**Result/conclusion**] The theoretical framework and implementation method are helpful for accurate information recommendation, improving patient experience and improving decision-making quality, and provide theoretical and practical guidance for construction and application of OHC-UP.

Keywords: user persona online health community diabetes